IMPROVING EFFICIENCY OF ONTOLOGY MAPPING IN SEMANTIC WEB USING CUT ARC ALGORITHM

S.Raja Ranganathan[1] Dr.M.Marikkannan [2] Dr.S.Karthik [3]

[1]Assistant Professor, Dept of CSE, SNS College of Technology-Coimbatore.
E-Mail ID: rajaranganathan.s@gmail.com
[2]Assistant Professor, Dept of CSE, Instuite of Road and Transport Technology-Erode.
E-Mail ID : mmk.irtt@gmail.com
[3]Professor and Dean, Dept of CSE, SNS College of Technology-Coimbatore.
E-Mail ID : profskarthik@gmail.com

Abstract-The World Wide Web (WWW) is a huge collection of information this allows the people to share information from the repositories of databases which are available globally. From day to day the information which is available inside database is growing rapidly also the number of user interaction with World Wide Web also increased. The search engines play a major role to extract these information to the user, there are many search engines available today but retrieving the most optimized result is a difficult task, however to overcome the problems in search engines for retrieving the optimized results for user queries, Semantic search technology plays a vital role. The main purpose of the Semantic Web is driving the evolution of the current Web by enabling users to find, share, and combine information more easily. In this paper we briefly review about new semantic architecture and propose a new sub graph algorithm for relation based semantic web technology this helps to improve the efficiency and consistency of semantic web.

Index Terms—Artificial Intelligence, Extensible Markup Language, Ontology Web Language, Resource Descriptive Framework, Resource Descriptive Framework Structure, Uniform Resource Locater, World Wide Web.

1 INTRODUCTION

The Web Search Engines methodologies, following recent years is always less uncommon that the results provide by them are greater burden of useless pages to the users. The third-generation Web architecture the semantic web, provides the three layered architecture possibly allowing overcoming the limitation and burdens that caused in existing searching methodology. There are several search engines have been proposed and adopted, their main contribution is increasing information retrieval accuracy by exploiting a key content of Semantic Web resources, mainly based on relations between the concepts however, in order to rank or priorities results, most of the existing search solutions need to work on the whole annotated knowledge database. The semantic web has a strong architecture, which is capable of providing solution for all the problems of existing search methodology. In Semantic Web, each page or concepts will have semantic metadata concerning the Web page itself. Annotations are purely based on set of concepts and relations among those concepts, these annotations are very useful to produce the enhanced and most reliable results to the users.

One of the main components in semantic web is (RDF) Resource Description Framework a new standard of W3C the search efficiency has been improved by multiple Combinations made for user's keywords the RDF looks subject, Predicate and Object for each statement the user Intend to search. The RDF is purely an XML language and RDF enables exchange and reuse of structured metadata. The RDF coding given below

```
<rdf:Description rdf:ID="Coimbatore">
<has Accommodation rdf:resource="hotels">
</rdf:Description>
<rdf:description rdf:ID="hotel">
<has Accommodation rdf:resource="#three star">
</rdf:Description>
```

Fig-1 coding for RDF

The above XML coding will forms an RDF model, which would intend to search the hotels available in Coimbatore it would effectively form a metadata that helps to enhance the results of three star hotels available within the city. The second important component in semantic infrastructure is Ontology [2] this helps to make the relation among the successful concepts. The ontologies use OWL

Web ontology language in different levels we can express they are OWL Lite, OWL DL and OWL FULL ordering by increasing level. The Semantic Web will support more efficient routing, expertise decision, integration and reuse of data and provide support for interoperability problem which cannot be resolved with current web technologies. The main issue we address in

this paper is how to improve the efficiency of semantic web and how to make use of concept relations and to improve the relations between concepts to retrieve a more precise and smaller result.

2 LITRATURE SURVEY

The extraction of information from World Wide Web is not a new mechanism but we have to face challenges in information retrieval in many ways. There is different kind of search engines available in WWW [10] each search engine follows a unique mechanism of indexing and processes of search of its own so the information extraction as well as the result produced by these search engines are not the same. Some of the popular search engines such as GOOGLE, YAHOO, BING and ALTA VISTA produce results based on queries after the keyword are processed. They only search information available on the web page, recently updated, some research group's such as SWOOGLE start delivering results from their semantics based search engines, and however most of them are in their initial stages they face certain problems in matching ontology and combining keywords in RDF[12]. The major problems facing by the search engines are they not able to gather content whole indexing in entire internet.

In the survey it is found [2] that, ontology based lexical analysis have been made like synonyms antonyms, and homonyms between the keywords but not represented for the concepts this is used to expand the query results in a normal data form.in the above case the search engine only they considering is normal pattern based search not the semantic web search. The ontology based information retrieval [3][4][5] their work is on how the information's are retrieved from world wide web but not focus on semantic relations, which is the primary concept in semantic web[6][7]. The previous ranking strategies only work on the basis of score measuring the distance between the systematic description between the user query and the information retrieval resource, and computation taken place based on the ratio between relation instance of nodes and linking concepts.

Onto Look gives different solutions for to relate the pages with the graph based representation. Concepts and relations are modeled as vertices and weighted edge, then less relevant concepts that means nodes are removed fromthe graph. This allows the generation of candidate relation key word set [CRKS], to be submitted to the annotated knowledge database, which significantly reduce the presence of uninterested pages in the result set. The main drawback here is that there is no ranking strategy to be adopted for the results obtained to get prioritized. The page rank is important for the semantic web this would help to priorities the results in which published to the user, but the only renowned algorithm for web search is used by GOOGLE[6]

3 THE OVARALL SEMANTIC SYSTEM STRUCTURE

The semantic web architecture is dramatically advanced from the architectures that are followed from the existing architecture that used in existing information retrieval mechanism. Semantic data processing includes these earlier statistical and natural langue techniques, and enhances these with semantic processing tools. The semantic web is fully automated conversion of text and storage of unstructured data resources in a semantic web database. Semantic Web automatically extracts and processes the concept (pages) and information's mapped in the database in a range of highly flexible manner. The components of the system structures explained clearly in next sections

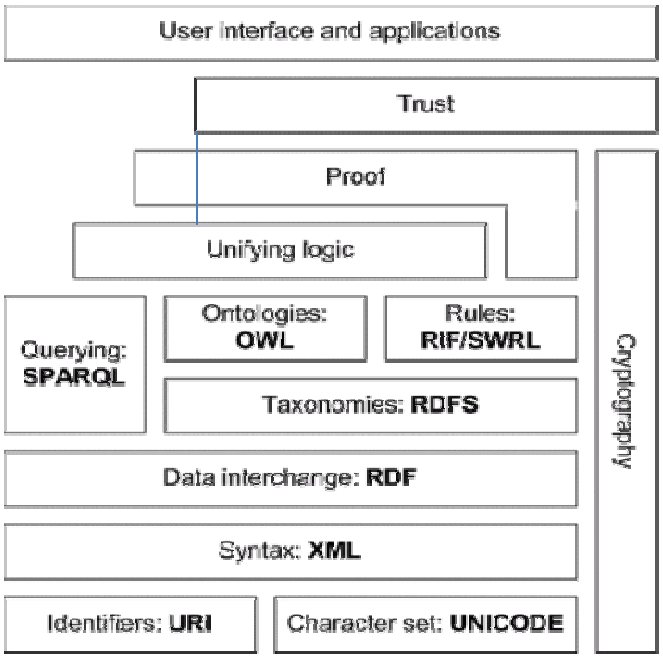


Fig 2.Semantic Structure

4 THE SEMANTIC-MICRO STRUCTURE

In semantic web architecture the core component is resource description framework (RDF). In the resource description framework all the information is expressed as a graph based expressions. The primary focus of RDF is to produce metadata relations of semantic resources,

such as the title of the pages, author who wrote the articles and modification date of a Web page dynamically, but it can also be used for storing any other data within knowledge database. Resource Description Framework a new standard of W3C the search efficiency has been improved by multiple combinations made for user's keywords the RDF looks subject, Predicate and Object for each statement the user intend to search. The RDF is purely an XML language and RDF enables exchange and reuse of structured metadata

```
<ontologypath>
<ontology=http://211.65.231.23/hotel/Cbe/ontology/travel.owl"/>
</ontologypath>
<rdfdescription>
<rdf:RDF
xmlns      =      "http://www.owl-ontologies.com/travel.owl#"
xml:base   =      "http://www.owl-ontologies.com/travel.owl#"
xmlns:cbe  =      "http://211.65.231.23/hotel/Cbe#"
xmlns:rdf  =      "http://www.w3.org/2012/12/owl#"
<city rdf:about="cbe:Coimbatore"/>
<hotel rdf:about="cbe:residency"/>
<hotel rdf:about="cbe:hotel"/>
<Accommodationrating rdf:about="cbe:fivestar/>
<activity rdf:about="cbe:sauna"/>
<activity rdf:about="cbe:bowling"/>
<activity rdf:about="cbe:swimming"/>
</rdf:RDF>
</rdfdescription>
```

Fig-3 RDF Coding to search a Hotel

As shown in Fig-3, the first segment, that is, the content between label <ontologypath> and </ontologypath>, indicates the present location that the ontology of the current (concepts) Web page belongs to. The ontology is capable to interpret the metadata in semantic annotation. The RDF coding shown above helps to find the hotel and its specifications in World Wide Web. The above mentioned RDF structure works as a key for web crawler based on the information above it will automatically extracts information from World Wide Web. We can include multiples of searching criteria within resource description framework (RDF). the RDF makes triple syntax arrangement and automatically analyze the keywords subject, predicate, and object based on this it crawl information from WWW.

5 SEMANTIC ARCHITECTURE AND WORK FLOW MACHANISM

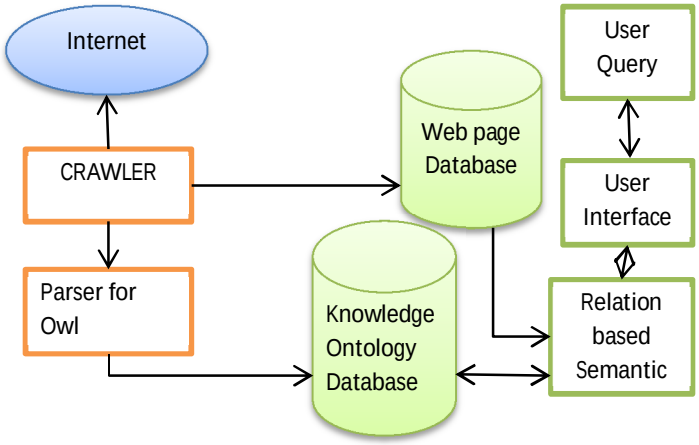


Fig-4 Semantic Web Architecture

The above diagram fig-4 shows the architecture of semantic web search engine. The crawler play a vital role in search engine, the primary work of crawler is to extract the key information from World Wide Web. The basic crawler search is shown in semantic microstructure Fig-3 with the semantic rule (RDF) and ontology, which is briefly described in OWL[4] (Web Ontology Language) document in internet. The extracted huge collection of web pages are exported to web page database and stored for the use of future retrieving URLs and corresponding Web pages. The work of OWL parser is to map the items inside the database. The primary reason for these mapping of items in relational database is I) the relation database and its items are mature and widely used II) until now, there is no single ontology query language recognized and the language used by any application support. The last stage of information collection by the crawler is the RDF label, which is the only formal method to annotate the web page contents. We also send the RDF [12] label to OWL parser for to develop a new table named "page description" within the relation knowledge database. Then the final result would be published to the users. The relation based search has the

interaction between the knowledge databases. The knowledge databases all the ontology network resources are mapped by graphs and sub graphs as shown in Fig 6 and 7. RDFS and OWL [4] have semantics defined procedures and this semantics can be used for reasoning within ontologies and knowledge bases described using these languages. To provide rules beyond the constructs available from these languages, rule semantic based languages are being standardized for the semantic web as well

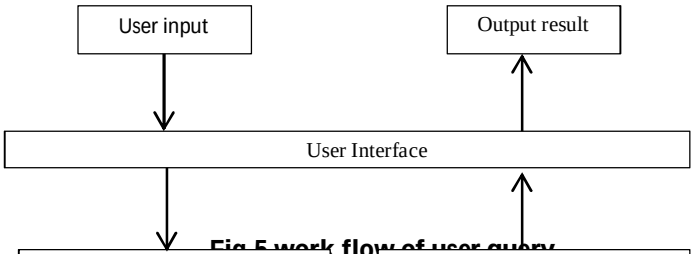


Fig-5 work flow of user query

semantic search engine ("onto look") the first stage "onto look" will analyze the query forms the keyword as multiple

"travel source" and "travel destination" our system will

combine the pages or concepts to combined concept pairs and these pairs are send to the ontology based knowledge database to retrieve all the relations produced by ontology between combined concept pairs. After all relations between combined concept pairs are retrieved from the ontology knowledge database, we can form a concept onto-relation graph [8] based on these relations and concepts. The ellipse shown in the diagram denotes the concepts, some of the words near to the concepts denotes the keywords, arcs denotes the relation between two concepts, finally the number near to arc denotes the number of relation in ontology database. as shown in Fig-6 for examples the upper ellipse represent the concept "destination" and "accommodation", respectively. And, the number "4" near the arc between the two concepts represents that there are three relations in the ontology database. In the next section, we will formally describe the concept-relation graph

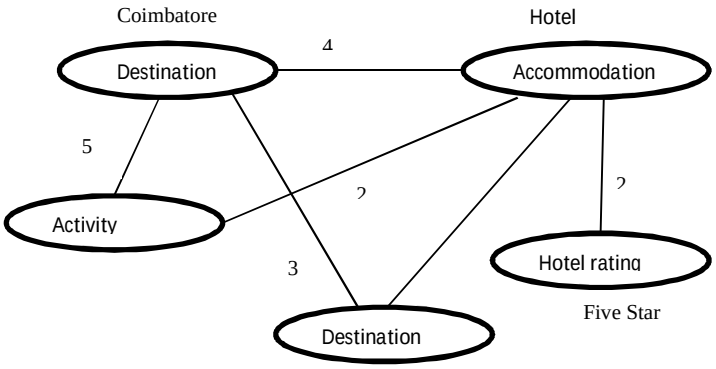
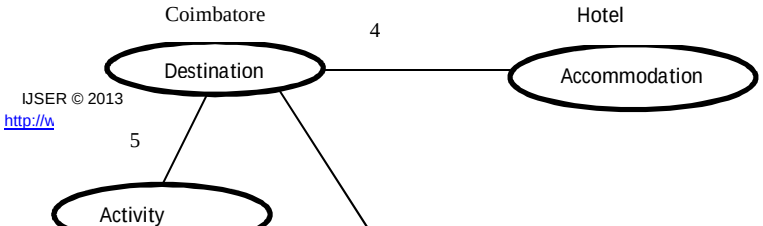


Fig-6 The graph based representation ontology

From the main ontology graph [9] a sub graph Fig-7 is derived for to improve the performance of information retrieval. In the next step, the ontology relations will cut some arcs from the graph and form some sub graphs for to make easy mapping performance, In each sub graph, there are some denotable relations between the concepts. The larger the number near the arc is, the more relations between the concepts exist. The number zero means a pure keyword search without any relationship appended in ontology mapping. Because there are some relations between the keywords which users input, the result set retrieved from the database will be close to the users' intention when less-ranked arcs were cut from the graph. Finally, the system fetches the relation and its corresponding keyword pair from each arc in sub graphs to form a property-keyword candidate set. Then, the property keyword candidate set is sent to the database to get a retrieved result set for the users.



6 EFFICIENT ALGORITHM FOR GENERATING SUB GRAPH

The relation based sub graph G is to find and label the arc to be cut in the arc set R. For example, if we are going to cut three arcs, it means we have to find all the sets whose length equals 2 in the power set 2^r of R. the superscript r denotes the number of relation between items. Fig. 8 is the going representation of the relation based sub graph algorithm. From the definition of power set, we can determine that the time complexity of the cut arc algorithm is. It seems to be not suitably fitted in practice of some cases, but, fortunately, few people will submit large numbers of keywords to search engines.

```
// procedure for sub graph
// sub-vec is a vector,  stores the graph sequence
vector sub-vec;
subvec(){
 // begin with one graph arc, we cut each arc
 //every time while one added each time
          for(i=1;i<=|sub-vec|;i++)
          getSubset(int n,vector sub-vec);}
//Procedure getsubset
//n is the number arc thet is to be cut
getsubsetgraph(int n){
//define the stack that temporarly store the arc that is to be
taken out
Stack graphstackk;
          for(int i=0;i<=|sub-vec|;i++){
          j=i;
          graphstack.push(sub-vec[j])
          while(!graphstack.empty()){
          if(graphstack.size()==n){
//Set of arc is the number of src stack we obtained
//the process can be start for this arc
process_sub-graph();
// the stack leads to popup process and ready for next arc to
push
graphstack.popup();
}
else
{
if(j++!=sub-vec.size())
//position to the pointer j do not reach the top arc
//then push item into stack
graphstack.push(sub-vec[j[);
else
{
//if pointer reach top then get the position of arc top
j=getposition(graphstack.gettop());
graphstack.popup();
}}}}}
```

Fig-8 Sub graph cut arc algorithm

7 PERFORMANCE MEASURES

TABLE-1

Construction made without sub graph algorithm

Number of Keywords	Relation between keywords	Number of sub graph processed	Number of property relation produced	CPU processing time(millisecconds)
1	0	0	0	0
2	2	2	5	37
3	3	4	15	73
4	8	56	234	182
5	9	143	32454	2354

The initial analysis in table-1 has been made without use of cut arc sub graph algorithm which is mentioned in previous sections totally five keywords being searched. The relation obtained for five keywords is nine as well as number of sub graph processed without cut arc methodology is above hundred, in this cases most of the subgraph are revisiting ones or unwanted relation subgraph generated. Number of property relation between the concepts or keywords also high in this sense there is a huge revisiting of same nodes that visited earlier. The next table shows all data obtained with the use of efficient sub graph algorithm while compare to the previous table the processing time of the CPU is 65 percent decreased, hence the user will able to get the most efficient result in more quicker time. In the previous table the most of CPU processing time is utilized by the unwanted relation made between nodes and by the revisiting sub graphs generated.

The problem is efficiently handled by the cut arc subgraph algorithm the subgraph processed with algorithm is efficient only the best nodes are selected and form more reliable relation within the concepts without any revisiting, mislead or blind links.

TABLE-2

Construction made with sub graph algorithm

Number of Keywords	Relation between keywords	Number of subgraph processed	Number of property relation produced	CPU processing time(millis econds)
1	0	0	0	0
2	2	2	5	36
3	3	4	15	58
4	8	36	183	135
5	9	78	1242	532

8 CONCLUSION

The semantic web is capable of extracting millions of information from World Wide Web. Handling that information is a huge task for Semantic search engine developers. In this paper we propose a new sub graph based cut arc algorithm that must improve the efficiency of information retrieval. Also this algorithm helps to eliminate the unwanted nodes that improve the search accuracy. Further work has been involved in improving the environment of Micro semantic Web with enhanced algorithms. Because of decentralized as well as heterogeneous structure of web pages it is impossible to use same ontology for all domains hence brief study of semantic communication is needed.

REFERENCES

[1] T. Berners-Lee and M. Fischetti, "Weaving the Web," Chapter Machines and the Web, pp. 177-198, 1999.

[2] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web,"Scientific Am., vol. 284, no. 5, pp. 34-43, 2001.

[3] D. Beckett, "RDF/XML Syntax Specification (Revised),"http:// www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/, 1994.

[4] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D.L. McGuinness, P.F. Patel-Schneider, and L.A. Stein, "OWL Web Ontology Language Reference," http://www.w3.org/TR/2004/ REC-owl-ref-20040210/, 2004.

[5] A. Go´mez-Pe´rez and O. Corcho, "Ontology Languages for the Semantic Web," IEEE Intelligent Systems, vol. 17, no. 1, pp. 54-60, Jan.-Feb. 2002.

[6] The Google.com search engine, http://www.google.com/, 2004.

[7] L. Ding, T. Finin, A. Joshi, R. Pan, R.S. Cost, Y. Peng, P. Reddivari, V. Doshi, and J. Sachs, "Swoogle: A Search and Metadata Engine for the Semantic Web," Proc. 13th ACM Int'l Conf. Information andKnowledge Management (CIKM '04), pp. 652-659, 2004.

[8]. L. Ding, T. Finin, A. Joshi, Y. Peng, R. Pan, and P. Reddivari, "Search on the Semantic Web," Computer,vol. 38, no. 10, pp. 62-69, Oct. 2005.

[9] L. Ding, P. Kolari, Z. Ding, and S. Avancha, "Using Ontologies in the Semantic Web: A Survey,"Ontologies, pp. 79-113, Springer, 2007

[10]. R. Guha, R. McCool, and E. Miller, "Semantic Search," Proc. 12th Int'l Conf. World Wide Web (WWW'03), pp. 700-709, 2003.

[11] Z. Gyongyi and H. Garcia-Molina, "Spam: It's Not Just for Inboxes Anymore," Computer, vol. 38, no. 10,pp. 28-34,

[12] C. Junghoo, H. Garcia-Molina, and L. Page, "Efficient Crawling through URL Ordering," Computer Networks and ISDN Systems, vol. 30, no. 1, pp. 161-172, 1998

S.Raja Ranganathan has completed his MCA Degree in Computer Science from SNSCT, in Anna University, Chennai in year 2008.He obtained his Master's Degree in Computer Science and Engineering from SNSCT in Anna University, Coimbatore in year 2010. His research interest focuses on Web mining and Semantic web. He has presented more than three papers in national conferences. At present, He is an Assistant Professor of Computer Science and Engineering in SNS College of Technology,

Anna University, Coimbatore. He is pursuing part time Ph.D. in Anna University of technology Coimbatore.

Dr.M.Marikkannan received the B.Eng. degree in computer science and engineering from the Government College of Engineering,Tirunelveli, India in 1994, and M.Eng.degree in computer science and engineering from College of Engineering, Anna University,Chennai, India in 1999. Currently, he is an Assistant professor at the Department of Computer Science and Engineering, Institute of Road and Transport Technology (IRTT),Erode, India.His research interests include temporal database management systems and object oriented systems.

Professor Dr.S.Karthik is presently Professor & Dean in the Department of Computer Science & Engineering, SNS College of Technology, affiliated to Anna University-Coimbatore, Tamilnadu, India. He received the M.E degree from the Anna University Chennai and Ph.D degree from Ann University of Technology, Coimbatore. His research interests include network security, web services and wireless systems. In particular, he is currently working in a research group developing new Internet security architectures and active defense systems against DDoS attacks. Dr.S.Karthik published more than 35 papers in refereed international journals and 25 papers in conferences and has been involved many international conferences as Technical Chair and tutorial presenter. He is an active member of IEEE, ISTE, IAENG, IACSIT and Indian Computer Society.